

Optimal Estimation of High-Dimensional Covariance Matrices with Missing and Noisy Data

Meiyin Wang, Wanzhou Ye

Department of Mathematics, College of Science, Shanghai University, Shanghai, China
Email: wangmeiyin9@163.com, wzhy@shu.edu.cn

How to cite this paper: Wang, M.Y. and Ye, W.Z. (2024) Optimal Estimation of High-Dimensional Covariance Matrices with Missing and Noisy Data. *Advances in Pure Mathematics*, 14, 214-227.

<https://doi.org/10.4236/apm.2024.144013>

Received: March 14, 2024

Accepted: April 12, 2024

Published: April 15, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The estimation of covariance matrices is very important in many fields, such as statistics. In real applications, data are frequently influenced by high dimensions and noise. However, most relevant studies are based on complete data. This paper studies the optimal estimation of high-dimensional covariance matrices based on missing and noisy sample under the l_1 norm. First, the model with sub-Gaussian additive noise is presented. The generalized sample covariance is then modified to define a hard thresholding estimator $\hat{\Sigma}_F^{th}$, and the minimax upper bound is derived. After that, the minimax lower bound is derived, and it is concluded that the estimator presented in this article is rate-optimal. Finally, numerical simulation analysis is performed. The result shows that for missing samples with sub-Gaussian noise, if the true covariance matrix is sparse, the hard thresholding estimator $\hat{\Sigma}_F^{th}$ outperforms the traditional estimate method.

Keywords

High-Dimensional Covariance Matrix, Missing Data, Sub-Gaussian Noise, Optimal Estimation

1. Introduction

The covariance matrix is a key component in various fields, particularly statistics. However, when dealing with many statistical situations, the covariance matrix is usually unknown. As a result, estimating the covariance matrix is extremely significant, and it is frequently utilized in signal processing, genomics, financial mathematics, and other domains. When the dimension p is fixed and

the sample size n is sufficiently large, the sample covariance matrix is commonly used to estimate the true covariance matrix. However, with advancements in information technology and various other technologies, there is a growing challenge in estimating large covariance matrices. Issues such as dimensionality and noise can significantly impact the effectiveness of using the sample covariance matrix to estimate the true covariance matrix. Moreover, in the era of big data, missing data is a common occurrence, making research on the estimation of high-dimensional covariance matrices based on missing and noisy data essential.

Bickel and Levina [1] proposed thresholding as a commonly used method for estimating high-dimensional covariance matrices and proved its consistency under the operator norm. However, there was no discussion on its optimality. Cai and Zhou studied the optimal estimation of sparse covariance matrices under the operator norm and Bregman divergence loss. They also proved that the thresholding estimator can achieve the optimal convergence rate under the spectral norm, see [2]. Cai and Zhou [3] provided the optimal estimation of the sparse covariance matrices under the l_1 norm loss. The thresholding described above is also referred to as hard thresholding, and its counterpart is soft thresholding [4] [5]. On this basis, Rothman, Levina, and Zhu [6] proposed generalized thresholding and proved its consistency. Cai and Liu [7] proposed adaptive thresholding. The adaptive estimation of high-dimensional sparse precision matrices was studied by Cai, Liu, and Zhou [8]. For bandable covariance matrices, [3], [9], and [10] conducted in-depth research.

In the case of missing data, Cai and Zhang [11] assumed that the missingness was independent of the data values and studied the optimal estimation of two classes of covariance matrices. Qi [12] explored the optimal estimation of sparse covariance matrices under the l_1 norm and the Frobenius norm, respectively. In addition, the lower bound for estimating bandable covariance matrices under the spectral norm was studied based on noisy and missing data, but its optimality is not considered. Shi [13] studied the optimal estimation of bandable covariance matrices based on missing and noisy sample data.

It is not difficult to find that the research on estimating high-dimensional covariance matrices is primarily based on complete data. However, the correlational research on missing data and noisy models remains critical. The articles listed above served as a tremendous source of inspiration for this paper's study topic and methods. This paper will provide corresponding research for the aforementioned issues. Sparse covariance matrices are widely employed in a variety of applications, including genomics. As a result, it is necessary to investigate the estimate of this kind of covariance matrix. The research in this paper can help people better estimate the high-dimensional covariance matrix when the sample is noisy and missing. Thus, it is convenient for many fields to better use high-dimensional data to obtain more useful information, and this paper provides them with a reliable theoretical basis.

The remaining sections of this paper are as follows: Section 2 will provide the

associated concepts and knowledge of covariance matrix estimation, which serves as the theoretical foundation for the research. In Section 3, we will study the optimal estimation of sparse covariance matrices with missing and noisy data. In Section 4, numerical simulation experiment will be performed to investigate the estimating effect of the estimator presented in Section 3. The fifth section summarizes the research content and discusses existing problems.

2. Theoretical Basis

This paper will primarily study the optimal estimation of covariance matrices under the l_1 norm. For a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, $\sigma(A) = \sqrt{\lambda(A^H A)}$ represents the singular values of A , while $\lambda(A^H A)$ represents the eigenvalues of $A^H A$. The operator norm of A is defined as

$$\|A\|_a = \max_{\|x\|_a=1, x \in \mathbb{R}^n} \|Ax\|_a = \max_{x \neq 0, x \in \mathbb{R}^n} \frac{\|Ax\|_a}{\|x\|_a}$$

There are three common operator norms:

- (1) l_1 norm: $\|A\|_{l_1} = \|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$;
- (2) spectral norm: $\|A\|_{sp} = \|A\|_2 = \sigma_1(A) = \max_i \sigma_i(A)$;
- (3) l_∞ norm: $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$.

Next, we will introduce the sub-Gaussian random vector. If there is a parameter $k > 0$ such that $\mathbb{E}(e^{sX}) \leq e^{k^2 s^2 / 2}, s \in \mathbb{R}$, the random variable X is considered a sub-Gaussian random variable with parameter k , that is, $X \sim Sub(k)$. It is easy to know that sub-Gaussian random variables include Gaussian random variables whose mean is 0 and all bounded random variables with a mean of 0. Assuming the random variable X is sub-Gaussian, its sub-Gaussian norm is denoted by

$$\|X\|_{\psi_2} = \sup_{p \geq 1} p^{-1/2} (\mathbb{E}|X|^p)^{1/p}.$$

A p -dimensional random vector $X = (X_1, X_2, \dots, X_p)^T$ is called the sub-Gaussian random vector if any linear combination of X_1, X_2, \dots, X_p is sub-Gaussian. That is, when $\tau > 0$, for any $t > 0$, $v \in \mathbb{R}^p$, and $\|v\|_2 = 1$, there is

$$P\{|v^T (X - \mathbb{E}X)| > t\} \leq e^{-t^2 / 2\tau}.$$

Assume that a p -dimensional random vector $X \in \mathbb{R}^p$ has the mean μ and the covariance matrix Σ . Covariance matrix estimation is the process of computing a covariance matrix $\hat{\Sigma}$ based on n independent copies $X_1, X_2, \dots, X_n \in \mathbb{R}^p$ of X and then using $\hat{\Sigma}$ to estimate Σ , i.e., making $\hat{\Sigma}$ approximate Σ in a certain sense. In this paper, minimax risk is used as a standard to measure the estimation effect. Suppose X_1 has a certain class of covariance matrices, and \mathcal{A} is a specific collection of X_1 's distributions. Then, under the specified matrix norm $\|\cdot\|$, [3] defines the minimax risk of estimating Σ over \mathcal{A} as

$$\mathcal{R}(\mathcal{A}) = \inf_{\hat{\Sigma}} \sup_{\mathcal{A}} \mathbb{E} \|\hat{\Sigma} - \Sigma\|^2.$$

When the vector's dimension p is smaller than the sample size n , the sample covariance matrix is typically utilized to estimate the true covariance matrix. The sample mean is

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

and the sample covariance matrix is

$$\hat{\Sigma} = (\hat{\sigma}_{ij})_{p \times p} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T. \quad (1)$$

However, as noted in Section 1, when the dimension p is substantially larger than the sample size n , utilizing the sample covariance matrix for estimating the true covariance matrix becomes inadequate. Based on the work of Cai et al., this paper will study the optimal estimation of high-dimensional covariance matrices based on missing and noisy data.

The missing completely at random (MCR) model is introduced below. MCR indicates that the missingness was random and independent of the data values. Suppose $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ is complete random sample from \mathbf{X} . Introducing vector $\mathbf{S}_k = \{0, 1\}^p, k = 1, 2, \dots, n$ as the observation index for \mathbf{X}_k , then

$$S_{jk} = \begin{cases} 1, & X_{jk} \text{ is observed,} \\ 0, & X_{jk} \text{ is missing.} \end{cases}$$

S_{jk} and X_{jk} represent the j th coordinate of vectors \mathbf{X}_k and \mathbf{S}_k , respectively.

We denote $\mathbf{X}^* = \{\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_n^*\}$ as the sample with missing data, where $\mathbf{X}_i^* = (X_{1i}S_{1i}, X_{2i}S_{2i}, \dots, X_{pi}S_{pi})^T$ is the i th observation sample. Additionally, define

$$n_{ij}^* := \sum_{k=1}^n S_{ik}S_{jk}, \quad 1 \leq i, j \leq p.$$

When $S_{ik}S_{jk} = 1$, the i th and j th components of vector \mathbf{X}_k^* are observed simultaneously, whereas $S_{ik}S_{jk} = 0$ indicates that they were not observed simultaneously. Thus, n_{ij}^* denotes the number of times the i th and j th components of \mathbf{X}^* are simultaneously observed. For convenience, let's define $n_i^* = n_{ii}^*$, $n_{\min}^* = \min_{i,j} n_{ij}^*$. It is simple to know that $n_{\min}^* \leq n_{ij}^* \leq \min\{n_i^*, n_j^*\}$. When the sample data are complete, $n_{ij}^* \equiv n$.

For sample \mathbf{X}^* with missing data, we substitute the generalized sample mean and generalized sample covariance matrix for the traditional sample mean and covariance matrix. The generalized sample mean is defined as the following:

$$\bar{\mathbf{X}}^* := (\bar{X}_i^*)_{1 \leq i \leq p}, \quad \bar{X}_i^* = \frac{1}{n_i^*} \sum_{k=1}^n X_{ik}^* = \frac{1}{n_i^*} \sum_{k=1}^n X_{ik} S_{ik},$$

the generalized sample covariance matrix is defined as

$$\hat{\Sigma}^* := (\hat{\sigma}_{ij}^*)_{1 \leq i, j \leq p}, \quad \hat{\sigma}_{ij}^* = \frac{1}{n_{ij}^*} \sum_{k=1}^n (X_{ik} - \bar{X}_i^*)(X_{jk} - \bar{X}_j^*) S_{ik} S_{jk}. \quad (2)$$

3. Covariance Matrix estimation

This paper assumes that the covariance matrix is sparse, which means that the majority of its components are 0 or insignificant, and the distribution of non-zero elements is irregular. First, we introduce the parameter space $\mathcal{G}_q(\rho, c_{n,p})$ of the sparse covariance matrices:

$$\mathcal{G}_q(\rho, c_{n,p}) = \left\{ \Sigma = (\sigma_{ij})_{1 \leq i, j \leq p} : \max_{1 \leq j \leq p} \left\{ |\sigma_{[k]j}|^q \right\} \leq \frac{c_{n,p}}{k}, \forall k, \max \sigma_{ii} \leq \rho \right\},$$

where $0 \leq q < 1$, and $\sigma_{[k]j}$ represents the element with the k th largest absolute value in the j th column of matrix Σ . When $q = 0$, each column of the matrices in $\mathcal{G}_q(\rho, c_{n,p})$ has at most $c_{n,p}$ non-zero components, usually assuming $c_{n,p} \geq 1$.

3.1. Noisy Model

Assuming the complete random vector $\mathbf{X} \in \mathbb{R}^p$ has the covariance matrix $\Sigma_X = (\hat{\sigma}_{ij})_{p \times p}$. Using a p -dimensional random vector \mathbf{F} to represent noisy data, the noisy model can be expressed as

$$\mathbf{F} = \mathbf{X} + \boldsymbol{\varepsilon}, \quad (3)$$

where $\boldsymbol{\varepsilon} \in \mathbb{R}^p$ represents noise. In this section, we hope to build a $p \times p$ matrix $\hat{\Sigma}_F$ based on n independent random noisy samples $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_n \in \mathbb{R}^p$ of \mathbf{F} . We next use $\hat{\Sigma}_F$ to estimate the covariance matrix Σ_X of the random vector \mathbf{X} .

The noisy sample with missing data are represented by $\mathbf{F}^* = \{\mathbf{F}_1^*, \mathbf{F}_2^*, \dots, \mathbf{F}_n^*\}$, where $\mathbf{F}_i^* = (F_{1i} S_{1i}, F_{2i} S_{2i}, \dots, F_{pi} S_{pi})^T$ is the i th observation sample. The definition of the generalized sample mean is as follows:

$$\bar{\mathbf{F}}^* := (\bar{F}_i^*)_{1 \leq i \leq p}, \quad \bar{F}_i^* = \frac{1}{n_i^*} \sum_{k=1}^n F_{ik}^* = \frac{1}{n_i^*} \sum_{k=1}^n F_{ik} S_{ik},$$

the generalized sample covariance matrix is defined as

$$\hat{\Sigma}_F^* := (\hat{\sigma}_{ij}^*(\mathbf{F}))_{1 \leq i, j \leq p}, \quad \hat{\sigma}_{ij}^*(\mathbf{F}) = \frac{1}{n_{ij}^*} \sum_{k=1}^n (F_{ik} - \bar{F}_i^*)(F_{jk} - \bar{F}_j^*) S_{ik} S_{jk}. \quad (4)$$

Two new assumptions in [12] are presented below.

Assumption 1. The observation index $\mathbf{S} := \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_n\}$ can be random or deterministic, but it is independent of the noisy observation sample $\mathbf{F} := \{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_n\}$.

Assumption 2. The random vectors $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_n$ are i.i.d., where $\mathbf{F}_k = \mathbf{X}_k + \boldsymbol{\varepsilon}_k$, and

$$\mathbf{X}_k = \boldsymbol{\Gamma} \mathbf{Z}_k + \boldsymbol{\mu}, \quad \boldsymbol{\varepsilon}_k = \boldsymbol{\Gamma}^\varepsilon \mathbf{Z}_k, \quad k = 1, 2, \dots, n.$$

$\boldsymbol{\mu}$ represents a fixed p -dimensional mean vector. $\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^\varepsilon \in \mathbb{R}^{p \times q}$ ($p \leq q$) are

fixed matrices with $\mathbf{I}\mathbf{T}^T = \mathbf{\Sigma}$ and $\mathbf{T}^\varepsilon \mathbf{T}^{\varepsilon T} = \mathbf{\Sigma}^\varepsilon$. Each component of the random vector $\mathbf{Z}_k = (Z_{1k}, Z_{2k}, \dots, Z_{qk})^T$ i.i.d. sub-Gaussian with a variance of 1 and a mean of 0. For any $s > 0$, there exists a parameter $\tau > 0$ such that $\mathbb{E}(e^{sZ_k}) \leq e^{\tau s^2/2}$, that is, $Z_{ik} \sim \text{Sub}(\tau)$.

3.2. Upper Bound for Estimating Sparse Covariance Matrix

The hard thresholding estimator based on complete data was proposed by [1]. When most of the elements in each row or column of the true covariance matrix are close to zero or negligible, set the elements of the sample covariance matrix below a certain threshold to 0, and leave the remaining elements unaltered to estimate the true covariance matrix, so as to reduce the error. In [2],

$$P\left\{|\hat{\sigma}_{ij} - \sigma_{ij}| \leq t\right\} \geq 1 - Cp^{-8}$$

for $t = \gamma\sqrt{\log p/n}$, where C is a constant. The threshold is set to $\gamma\sqrt{\log p/n}$.

In this paper, it is extended to the case of missing and noisy data. According to Lemma 4.6 in [12], if Assumption 1 and Assumption 2 are both hold, then there are two absolute constants C and c greater than 0, such that

$$P\left\{|\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \leq x\right\} \geq 1 - C \exp\left\{-cn_{\min}^* \min\left(\frac{x^2}{\tau^4 \sigma_{ii} \sigma_{jj}}, \frac{x}{\tau^2 \sqrt{\sigma_{ii} \sigma_{jj}}}\right)\right\} \quad (5)$$

for any $x > 0$. Since $\sigma_{ii} \sigma_{jj} \leq \rho^2$, the above can be simplified to: there are constants $C > 0$ and $\gamma > 0$, such that

$$P\left\{|\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \leq x\right\} \geq 1 - C \exp\left(-\frac{8}{\gamma^2} n_{\min}^* x^2\right). \quad (6)$$

Where $x \leq \rho$, and the constants C and γ only depend on ρ . Note that Inequality (6) can be written as $P\left\{|\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \leq x\right\} \geq 1 - Cp^{-8}$ when $x = \gamma\sqrt{\log p/n_{\min}^*}$.

The hard thresholding estimator $\hat{\Sigma}_F^{th}$ of the covariance matrices $\Sigma_X \in \mathcal{G}_q(\rho, c_{n,p})$ is defined by transforming the generalized sample covariance $\hat{\sigma}_{ij}^*(\mathbf{F})$ in Equation (4),

$$\hat{\Sigma}_F^{th} = (\hat{\sigma}_{ij}^{th}(\mathbf{F}))_{p \times p} = (\hat{\sigma}_{ij}^*(\mathbf{F}) \cdot I(|\hat{\sigma}_{ij}^*(\mathbf{F})| \geq \lambda))_{p \times p}, \lambda = \gamma \sqrt{\frac{\log p}{n_{\min}^*}}, \quad (7)$$

where γ is a constant and $\gamma > 0$.

The following is Lemma 1, which plays an important role in studying the minimax upper bound. Lemma 1 generalizes Lemma 8 in [8] from complete to noisy and missing sample.

Lemma 1. Define event $A_{ij} := \left\{|\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| > 4 \min(|\sigma_{ij}|, \lambda = \gamma\sqrt{\log p/n_{\min}^*})\right\}$, then there is constant $c > 0$, which only depends on ρ , such that

$$P\{A_{ij}\} \leq 2cp^{-\frac{9}{2}}.$$

Proof : Firstly, define event $B_1 := \{|\hat{\sigma}_{ij}^{th}(\mathbf{F})| \geq \lambda\}$. It is easy to know that

$$\begin{aligned} B_1 &\subset \{|\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| \geq |\hat{\sigma}_{ij}^{th}(\mathbf{F})| - |\sigma_{ij}| \geq \lambda - |\sigma_{ij}|\}, \\ B_1^c &\subset \{|\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| \geq |\sigma_{ij}| - |\hat{\sigma}_{ij}^{th}(\mathbf{F})| > |\sigma_{ij}| - \lambda\}. \end{aligned} \tag{8}$$

According to the definition of $\hat{\sigma}_{ij}^{th}(\mathbf{F})$ in Equation (7),

$$|\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| = |\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| I(B_1) + |\sigma_{ij}| I(B_1^c). \tag{9}$$

Next, we will prove this lemma in different cases. It can be obtained by simple calculation:

$$P\left\{|\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| \leq 4 \min(|\sigma_{ij}|, \lambda)\right\} \begin{cases} \geq 1 - Cp^{-9/2}, & |\sigma_{ij}| < \lambda/4, \\ \geq 1 - Cp^{-8}, & \lambda/4 \leq |\sigma_{ij}| \leq 2\lambda, \\ \geq 1 - 2Cp^{-8} & |\sigma_{ij}| > 2\lambda. \end{cases}$$

Therefore, there exists a constant $c > 0$, such that

$$P\left\{|\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| > 4 \min(|\sigma_{ij}|, \lambda)\right\} \leq 2cp^{-9/2}.$$

Next, we can obtain the upper bound for estimating the sparse covariance matrices by utilizing the risk properties of thresholding estimator.

Theorem 1. If Assumption 1 and Assumption 2 hold, $\log p = o(\sqrt{n_{\min}^*})$ and $p \geq \sqrt{n_{\min}^*}$, then there is a constant $C > 0$ such that the hard thresholding estimator $\hat{\Sigma}_F^{th}$ defined by Equation (7) satisfies

$$\sup_{\Sigma_X \in \mathcal{G}_q(\rho, c_{n,p})} \mathbb{E} \|\hat{\Sigma}_F^{th} - \Sigma_X\|_1^2 \leq Cc_{n,p}^2 \left(\frac{\log p}{n_{\min}^*}\right)^{1-q}. \tag{10}$$

Proof: Easy to know, $\|\hat{\Sigma}_F^{th} - \Sigma_X\|_1^2 = \left[\max_j \sum_{i=1}^p |\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}|\right]^2$. If event A_{ij} occurs,

$$\sum_{i=1}^p |\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| > \sum_{i=1}^p 4 \min(|\sigma_{ij}|, \lambda).$$

Simple calculations show that

$$\sum_{i=1}^p \min(|\sigma_{ij}|, \lambda) = \left(\sum_{i \leq k'} + \sum_{i > k'}\right) \min(|\sigma_{ij}|, \lambda) \leq \left(\sum_{i \leq k'} + \sum_{i > k'}\right) \min\left(|\sigma_{[i]j}|, \gamma \sqrt{\frac{\log p}{n_{\min}^*}}\right).$$

According to the definition of $\mathcal{G}_q(\rho, c_{n,p})$, we know that $\max_{1 \leq j \leq p} \left\{|\sigma_{[i]j}|\right\}^q \leq c_{n,p}/i$, so $|\sigma_{[i]j}| \leq (c_{n,p}/i)^{1/q}$. Select the constant k' to satisfy $k' = \left\lfloor c_{n,p} (n_{\min}^*/\log p)^{q/2} \right\rfloor$, so

$$\begin{aligned} \sum_{i=1}^p \min(|\sigma_{ij}|, \lambda) &\leq \sum_{i \leq k'} \min\left(\left(\frac{c_{n,p}}{i}\right)^{1/q}, \gamma \sqrt{\frac{\log p}{n_{\min}^*}}\right) + \sum_{i > k'} \min\left(\left(\frac{c_{n,p}}{i}\right)^{1/q}, \gamma \sqrt{\frac{\log p}{n_{\min}^*}}\right) \\ &\leq k' \gamma \sqrt{\frac{\log p}{n_{\min}^*}} + \sum_{i > k'} \left(\frac{c_{n,p}}{i}\right)^{1/q} \leq C_1 c_{n,p} \left(\frac{\log p}{n_{\min}^*}\right)^{1-2/q}. \end{aligned}$$

Let the matrix $\mathbf{D} = (d_{ij})_{p \times p}$ satisfy $d_{ij} = |\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}| \mathbf{I}(A_{ij})$, we have

$$\mathbb{E} \left\| \hat{\Sigma}_{\mathbf{F}}^{th} - \Sigma_{\mathbf{X}} \right\|_1^2 \leq 2\mathbb{E} \left\| \hat{\Sigma}_{\mathbf{F}}^{th} - \Sigma_{\mathbf{X}} - \mathbf{D} \right\|_1^2 + 2\mathbb{E} \|\mathbf{D}\|_1^2.$$

Then, it is straightforward to acquire

$$\begin{aligned} 2\mathbb{E} \left\| \hat{\Sigma}_{\mathbf{F}}^{th} - \Sigma_{\mathbf{X}} - \mathbf{D} \right\|_1^2 &= 2\mathbb{E} \left[\max_j \sum_{i=1}^p \left| \hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij} - (\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}) \mathbf{I}(A_{ij}) \right| \right]^2 \\ &\leq 2\mathbb{E} \left[\max_j \sum_{i=1}^p \left| (\hat{\sigma}_{ij}^{th}(\mathbf{F}) - \sigma_{ij}) \mathbf{I}(A_{ij}) \right| \right]^2 \leq C_2 c_{n,p}^2 \left(\frac{\log p}{n_{\min}^*} \right)^{1-q}. \end{aligned}$$

Therefore, we only need to prove that $2\mathbb{E} \|\mathbf{D}\|_1^2$ is negligible.

Firstly,

$$\begin{aligned} \mathbb{E} \|\mathbf{D}\|_1^2 &= \mathbb{E} \left[\max_j \sum_{i=1}^p |d_{ij}| \right]^2 \leq p^2 \mathbb{E} \left(\max_{i,j} d_{ij} \right)^2 \\ &= p^2 \mathbb{E} \left[\left(\max_{i,j} d_{ij}^2 \right) \mathbf{I}(A_{ij} \cap \{|\hat{\sigma}_{ij}^*(\mathbf{F})| \geq \lambda\}) \right] + p^2 \mathbb{E} \left[\left(\max_{i,j} d_{ij}^2 \right) \mathbf{I}(A_{ij} \cap \{|\hat{\sigma}_{ij}^*(\mathbf{F})| < \lambda\}) \right] \\ &\leq p^2 \mathbb{E} \left[\left(\max_{i,j} |\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \right)^2 \mathbf{I}(A_{ij}) \right] + p^2 \mathbb{E} \left[\left(\max_{i,j} |\sigma_{ij}| \right)^2 \mathbf{I}(A_{ij}) \right] \\ &= E_1 + E_2. \end{aligned}$$

According to the Cauchy-Schwartz inequality, we know that $|\sigma_{ij}|^2 \leq \sigma_{ii} \sigma_{jj}$, and because $P\{A_{ij}\} \leq 2cp^{-9/2}$, $p \geq \sqrt{n_{\min}^*}$, so

$$E_2 \leq p^2 \mathbb{E} \left(\max_{i,j} \sigma_{ii} \sigma_{jj} \right) P(A_{ij}) \leq p^2 \rho^2 2cp^{-9/2} \leq C_3 p^{-5/2} \leq \frac{C_4}{n_{\min}^*}.$$

In addition, $E_1 \leq p^2 \max_{i,j} \mathbb{E} |\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}|^2 P(A_{ij})$. From Inequality (5),

$$\begin{aligned} \mathbb{E} |\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}|^2 &= \int |\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}|^2 dP \leq \int_0^\infty xP\{|\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \geq x\} dx \\ &\leq \int_0^\rho xP\{|\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \geq x\} dx + \int_\rho^\infty xP\{|\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}| \geq x\} dx \\ &\leq \rho^2 + \int_\rho^\infty x \exp\left(-C_5 n_{\min}^* \frac{x}{\rho}\right) dx \leq \rho^2 + \frac{\rho^2}{C_5 n_{\min}^*} \exp(-C_5 n_{\min}^*). \end{aligned}$$

From Lemma 1, we have $P\{A_{ij}\} \leq 2cp^{-9/2}$ and $p \geq \sqrt{n_{\min}^*}$, so

$$\begin{aligned} E_1 &\leq p^2 \max_{i,j} \mathbb{E} |\hat{\sigma}_{ij}^*(\mathbf{F}) - \sigma_{ij}|^2 P(A_{ij}) \leq p^2 \left(\rho^2 + \frac{\rho^2}{C_5 n_{\min}^*} \exp(-C_5 n_{\min}^*) \right) 2cp^{-9/2} \\ &\leq p^2 \rho^2 2cp^{-9/2} + p^2 \frac{\rho^2}{C_5 n_{\min}^*} \exp(-C_5 n_{\min}^*) 2cp^{-9/2} \\ &\leq \frac{C_4}{n_{\min}^*} + \frac{C_6}{n_{\min}^*} \leq \frac{C_7}{n_{\min}^*} \end{aligned}$$

To sum up,

$$\mathbb{E} \|\mathbf{D}\|_1^2 \leq E_1 + E_2 \leq \frac{C_8}{n_{\min}^*} = O\left(\frac{1}{n_{\min}^*}\right).$$

3.3. Lower Bound for Estimating Sparse Covariance Matrix

Before studying the lower bound, introduce some useful lemmas and symbols.

Lemma 2. Assume P and V are two probability measures, with p and v representing their probability density functions. The total variation distance between P and V is $V(P, V) = 1 - \int \min(dP, dV)$. Define the total variation affinity as $\|P \wedge V\| := \int \min(dP, dV) = \int p(x) \wedge v(x) dx$. The Kullback divergence between P and V is expressed as $KL(P\|V) = \int p(x) \log[p(x)/v(x)] dx$. Thus, $\|P \wedge V\|$ and $KL(P\|V)$ satisfy the following inequality:

$$1 - \|P \wedge V\| \leq \sqrt{\frac{KL(P\|V)}{2}}. \tag{11}$$

Lemma 2 in [14] and Le Cam's lemma and its corollary in [2] [12] introduced below are important tools for proving minimax lower bound.

Lemma 3 (Le Cam). Suppose $\Theta = \{\theta_0, \theta_1, \dots, \theta_m\}$ is a finite set of parameters. Let L be a loss function and $l_{\min} := \min_{1 \leq i \leq m} \inf_t [L(t, \theta_0) + L(t, \theta_i)]$, then

$$\sup_{\theta \in \Theta} \mathbb{E} [L(\tilde{\theta}, \theta)] \geq \frac{1}{2} l_{\min} \|P_{\theta_0} \wedge \bar{P}\|.$$

$\tilde{\theta}$ is any estimator of θ based on the observed values of the probability measure $P_{\theta} (\theta \in \Theta)$, and $\bar{P} = \frac{1}{m} \sum_{i=1}^m P_{\theta_i}$.

Lemma 4. Suppose $\tilde{\Sigma}$ be any estimator of Σ_i based on the collection of probability measures $\{P_{\Sigma_0}, P_{\Sigma_1}, \dots, P_{\Sigma_m}\}$. We get

$$\sup_{1 \leq i \leq m} \mathbb{E} \|\tilde{\Sigma} - \Sigma_i\| \geq \frac{1}{2} \|P_{\Sigma_0} \wedge \bar{P}\| \cdot \inf_{1 \leq i \leq m} \|\Sigma_i - \Sigma_0\|,$$

where $\bar{P} = \frac{1}{m} \sum_{i=1}^m P_{\Sigma_i}$.

Before studying the minimax risk lower bound, it is advisable to construct a matrix with all off-diagonal elements equal to 0 except the first row or column. Let \mathcal{H} be the collection of $p \times p$ symmetric matrices in which exactly k non-diagonal elements in the first row or column equal to 1 and all other elements are 0. Let $k = \lfloor c_{n,p} (n_0 / \log p)^{q/2} \rfloor$. Define

$$\mathcal{G}_0 = \left\{ \Sigma = (\sigma_{ij})_{1 \leq i, j \leq p} : \Sigma = I_p \text{ or } \Sigma = I_p + a\mathbf{H}, \mathbf{H} \in \mathcal{H} \right\}, \tag{12}$$

where I_p represents the identity matrix of size $p \times p$, $a = \sqrt{\delta \log p / n_0}$, and δ is a constant. Assuming $\rho > 1$, $0 < \delta < \min\{1, 1/(4M)\}$, it is easy to know that $\mathcal{G}_0 \subset \mathcal{G}_q(\rho, c_{n,p})$.

Obtaining the lower bound requires two steps. Firstly, the subset of the parameter space constructed above is selected to simplify the proof. Secondly, calculate the total variation affinity between two probability measures.

Theorem 2. Let $1 \leq n_0 \leq n_{\min}^*$, $p \geq n_0^\nu (\nu > 1)$, and $\log p \leq n_0$. Assume

$c_{n,p} \leq M(n_0/\log p)^{(1-q)/2}$ with $0 \leq q < 1, M > 0$. For any $1 \leq n_0 \leq n_{\min}^*$, there exists a constant $c > 0$ such that the minimax risk lower bound for estimating the covariance matrix Σ_X satisfies

$$\inf_{\tilde{\Sigma}_F} \sup_{\Sigma_X \in \mathcal{G}_q(\rho, c_{n,p})} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_X\|_1^2 \geq cc_{n,p}^2 \left(\frac{\log p}{n_0}\right)^{1-q}. \tag{13}$$

where $\tilde{\Sigma}_F$ is any estimator of Σ_X based on noisy sample.

Proof: Assume $\mathcal{G}_0 = \{\Sigma_0, \Sigma_1, \dots, \Sigma_{m^*}\}$ has $m^* + 1$ elements, where Σ_0 represents the identity matrix and $\Sigma_i, i = 1, \dots, m^*$ represent the non-identity matrix, then $m^* = \text{Card}(\mathcal{G}_0) - 1 = C_{p-1}^k$.

Assume that $X_l, l = 1, \dots, n \stackrel{i.i.d.}{\sim} N(0, \Sigma_i), i = 1, \dots, m^*$, and the probability measure and probability density function are P_{Σ_i} and f_i , respectively, that is,

$\Sigma_X \in \mathcal{G}_0$. Let $F_l \stackrel{i.i.d.}{\sim} N(0, \Sigma_i(F))$ with $\Sigma_i(F) = \Sigma_i + s^2 I_p$ and $P_{\Sigma_i}(F)$ is the probability measure. Since $\mathcal{G}_0 \subset \mathcal{G}_q(\rho, c_{n,p})$, it is easy to know that

$$\inf_{\tilde{\Sigma}_F} \sup_{\Sigma_X \in \mathcal{G}_q(\rho, c_{n,p})} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_X\|_1^2 \geq \inf_{\tilde{\Sigma}_F} \sup_{\Sigma_X \in \mathcal{G}_0} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_X\|_1^2.$$

Therefore, to prove Inequality (13), just prove the following Inequality:

$$\inf_{\tilde{\Sigma}_F} \sup_{\Sigma_X \in \mathcal{G}_0} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_X\|_1^2 \geq cc_{n,p}^2 \left(\frac{\log p}{n_0}\right)^{1-q}. \tag{14}$$

Lemma 3.3 shows that

$$\sup_{\Sigma_X \in \mathcal{G}_0} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma\|_1 \geq \sup_{1 \leq i \leq m^*} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_i\|_1 \geq \frac{1}{2} \|P_{\Sigma_0}(F) \wedge \bar{P}(F)\| \cdot \inf_{1 \leq i \leq m^*} \|\Sigma_i(F) - \Sigma_0(F)\|_1.$$

Since $a = \sqrt{\delta \log p / n_0}$ and $k = \lfloor c_{n,p} (n_0 / \log p)^{q/2} \rfloor$, there exists a constant $c_1 > 0$ such that

$$\begin{aligned} \inf_{1 \leq i \leq m^*} \|\Sigma_i(F) - \Sigma_0(F)\|_1 &= \inf_{1 \leq i \leq m^*} \|(\Sigma_i + s^2 I_p) - (\Sigma_0 + s^2 I_p)\|_1 = \inf_{1 \leq i \leq m^*} \|aH\|_1 \\ &= ka \geq c_{n,p} \left(\frac{n_0}{\log p}\right)^{q/2} \cdot \sqrt{\delta \frac{\log p}{n_0}} \geq c_1 c_{n,p} \left(\frac{\log p}{n_0}\right)^{\frac{1-q}{2}} \end{aligned} \tag{15}$$

Obviously, to prove Inequality (14), we only need to prove that there is a constant $c_2 > 0$ such that $\|P_{\Sigma_0}(F) \wedge \bar{P}(F)\| \geq c_2$.

From $F_l \stackrel{i.i.d.}{\sim} N(0, \Sigma_i(F))$, we have

$$\text{KL}(P_{\Sigma_0}(F) \| P_{\Sigma_i}(F)) = \frac{1}{2} [\text{tr}(\Sigma_i^{-1}(F) \Sigma_0(F)) - \log \det(\Sigma_i^{-1}(F) \Sigma_0(F)) - p].$$

Let $B = -aH$, it is easy to know that $\Sigma_0(F) = \Sigma_i(F) + B$. Suppose the eigenvalues of $B \Sigma_i^{-1}(F)$ are ξ_1, \dots, ξ_p , then there are

$$\text{tr}(\Sigma_i^{-1}(F) \Sigma_0(F)) = \text{tr}[\Sigma_i^{-1}(F)(\Sigma_i(F) + B)] = \text{tr}[I_p + B \Sigma_i^{-1}(F)] = p + \sum_{i=1}^p \xi_i. \tag{16}$$

In addition, we can know that

$$\log \det(\Sigma_i^{-1}(\mathbf{F}) \Sigma_0(\mathbf{F})) = \sum_{i=1}^p \log(1 + \xi_i) = \text{tr}(\Sigma_i^{-1}(\mathbf{F}) \Sigma_0(\mathbf{F})) - p - \sum_{i=1}^p \frac{1}{2(1+\theta)} \xi_i^2, \tag{17}$$

where θ is a number between 0 and ξ_i . Putting Equation (16) and Equation (17) into $\text{KL}(P_{\Sigma_0}(\mathbf{F}) \| P_{\Sigma_i}(\mathbf{F}))$, we can get

$$\text{KL}(P_{\Sigma_0}(\mathbf{F}) \| P_{\Sigma_i}(\mathbf{F})) = \frac{1}{2} \left[p + \sum_{i=1}^p \frac{1}{2(1+\theta)} \xi_i^2 - p \right] = \frac{1}{2} \sum_{i=1}^p \frac{1}{2(1+\theta)} \xi_i^2 \leq \frac{1}{4} \sum_{i=1}^p \xi_i^2.$$

According to Theorem 1.3 in [12],

$$\sum_{i=1}^p \xi_i^2 = \text{tr}((\mathbf{B} \Sigma_i^{-1}(\mathbf{F}))^H \mathbf{B} \Sigma_i^{-1}(\mathbf{F})) = \|\mathbf{B} \Sigma_i^{-1}(\mathbf{F})\|_F^2 \leq \|\Sigma_i^{-1}(\mathbf{F})\|_{sp}^2 \|\mathbf{B}\|_F^2 \leq 2ka^2.$$

It is easy to see that $\bar{P}(\mathbf{F}) = 1/m^* \sum_{i=1}^{m^*} P_{\Sigma_i}(\mathbf{F})$, hence

$$\text{KL}(P_{\Sigma_0}(\mathbf{F}) \| \bar{P}(\mathbf{F})) \leq \frac{1}{m^*} \sum_{i=1}^{m^*} \text{KL}(P_{\Sigma_0}(\mathbf{F}) \| P_{\Sigma_i}(\mathbf{F})) \leq \frac{1}{8}.$$

Lemma 2 implies

$$\|P_{\Sigma_0}(\mathbf{F}) \wedge \bar{P}(\mathbf{F})\| \geq 1 - \sqrt{\frac{\text{KL}(P_{\Sigma_0}(\mathbf{F}) \| \bar{P}(\mathbf{F}))}{2}} = \frac{3}{4}.$$

That is, there exists a constant $c_2 > 0$ such that $\|P_{\Sigma_0}(\mathbf{F}) \wedge \bar{P}(\mathbf{F})\| \geq c_2$.

It is worth noting that Theorem 2 requires $c_{n,p} \leq M(n_0/\log p)^{(1-q)/2}$ ($M > 0$), which is a necessary condition. If $c_{n,p} > M(n_0/\log p)^{(1-q)/2}$, then

$$\inf_{\Sigma_F} \sup_{\Sigma_X \in \mathcal{G}_q(\rho, c_{n,p})} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_X\|_1^2 \geq \inf_{\Sigma_F} \sup_{\Sigma_X \in \mathcal{G}_q(\rho, M(n_0/\log p)^{(1-q)/2})} \mathbb{E} \|\tilde{\Sigma}_F - \Sigma_X\|_1^2 \gtrsim M^2.$$

Σ_X does not have a consistent estimator in this case.

Theorem 1 and Theorem 2 show that the estimator $\hat{\Sigma}_F^{th}$ we construct is rate-optimal over $\mathcal{G}_q(\rho, c_{n,p})$ under the l_1 norm.

4. Numerical Analysis

The optimal estimation of sparse covariance matrices based on missing and noisy data is derived in Section 3. This section compares the performance of the hard threshold estimator $\hat{\Sigma}_F^{th}$, as defined in Section 3, against the traditional estimator using numerical simulation.

Some symbols are presented before the numerical simulation begins. Assume the p -dimensional Gaussian random vector \mathbf{X} has a mean of $\boldsymbol{\mu}$ and a covariance matrix of Σ_X . n is the number of samples, and p is their dimension.

Here are the specific steps of numerical simulation.

1) Construct the sparse covariance matrix.

Assume $\boldsymbol{\mu}$ is a zero vector and Σ_X is a sparse matrix ($\Sigma_X \in \mathcal{G}_q(\rho, c_{n,p})$). Consult the construction of the sparse matrix in [11], let

$$\Sigma_X = \mathbf{I}_p + (\mathbf{B} + \mathbf{B}^T) / (\|\mathbf{B} + \mathbf{B}^T\|_1 + 0.01),$$

where $\mathbf{B} = (b_{ij})_{p \times p}$, and $P(b_{ij} = -1) = 0.1$, $P(b_{ij} = 0) = 0.8$, $P(b_{ij} = 1) = 0.1$.

2) Generate random samples according to the true covariance matrix.

After Σ_X is constructed, n p -dimensional random samples are first generated from the multivariate normal distribution with mean μ and covariance matrix Σ_X . The resulting n samples are then subjected to noise with a sub-Gaussian distribution, followed by random missing processing. This method produces sample data with missing and sub-Gaussian noise.

3) Compare the estimation effect of different estimators.

Based on the sample data with missing and sub-Gaussian noise, calculate the generalized sample covariance matrix $\hat{\Sigma}_F^*$ and the hard thresholding estimator $\hat{\Sigma}_F^{th}$ according to Equation (4) and Equation (7). Then compute the error between $\hat{\Sigma}_F^*$ and the real matrix Σ_X , as well as the error between $\hat{\Sigma}_F^{th}$ and the real matrix Σ_X , under the given norm.

After determining the values of n and p , repeat the above three steps 1), 2), and 3) 50 times, and take the mean value of the fifty error results as the standard for evaluating the estimation effect of different estimators in this case. The performance is better when the outcome is smaller. **Table 1** shows the experimental results.

The values of n and p are shown in the first two columns of **Table 1**. **Table 1** shows the average after 50 runs of the processes 1), 2), and 3) with n and p fixed. When the true covariance matrix Σ_X is sparse, the hard thresholding estimator $\hat{\Sigma}_F^{th}$ has a substantially better performance than the generalized sample covariance matrix $\hat{\Sigma}_F^*$ under any norm, especially when p is larger than n , that is, the dimension is high.

Therefore, when the dimension is very small in comparison to the sample size, the sample covariance matrix can be used to estimate the population covariance matrix. When estimating a high-dimensional sparse covariance matrix with sub-Gaussian additive noise and missing data, it is best to choose the hard thresholding estimator $\hat{\Sigma}_F^{th}$ given in Equation (7). This section provides some suggestions for application statisticians on how to select estimation methods.

Table 1. Results of estimating sparse covariance matrix.

n	p	$\ \hat{\Sigma}_F^* - \Sigma_X\ _1$	$\ \hat{\Sigma}_F^{th} - \Sigma_X\ _1$	$\ \hat{\Sigma}_F^* - \Sigma_X\ _{sp}$	$\ \hat{\Sigma}_F^{th} - \Sigma_X\ _{sp}$	$\ \hat{\Sigma}_F^* - \Sigma_X\ _F$	$\ \hat{\Sigma}_F^{th} - \Sigma_X\ _F$
50	20	7.5627	3.6435	3.9422	2.5361	7.3714	5.2233
	50	17.2511	4.8241	6.6115	3.1177	16.3977	8.3793
	100	34.0303	6.5240	10.7189	3.7975	31.6744	12.4028
	200	66.4459	7.3150	17.3097	4.1697	61.4820	17.3793
	300	99.1861	8.6805	23.5170	4.6641	92.1988	21.8822
	400	133.1350	9.9978	29.3351	4.9576	122.3446	25.6115
100	500	164.1220	10.1261	35.1697	5.0476	152.2275	28.4986
	50	12.1295	3.8520	4.7372	2.4652	12.7856	7.9800
	100	23.4557	4.8388	7.1643	2.8657	23.6691	11.5331
	200	44.8538	5.6213	10.9060	3.2010	45.0899	16.4487
	500	109.8831	6.9815	20.6747	3.6017	109.2537	26.2359

5. Summary and Outlook

In statistics and other fields, covariance matrix estimation is crucial. The estimation of high-dimensional covariance matrices has always been a hot topic with the rapid growth of numerous technologies.

Based on the missing and noisy sample data, this paper constructs a hard thresholding estimator $\hat{\Sigma}_F^{th}$, and studies its optimality. Section 3 shows that the hard thresholding estimator given in this paper is rate-optimal. The numerical simulation shown in Section 4 demonstrates that the hard thresholding estimator works well in situations where the true covariance matrix is sparse. When the true covariance matrix is not sparse, the estimation effect of the hard thresholding estimator has not been discussed.

This paper's research has limitations and areas that require more investigation:

- 1) This paper focuses solely on the optimal estimation of sparse covariance matrices based on noisy and missing data. More research is needed on the optimal estimation of other common high-dimensional covariance matrices.
- 2) If the sub-Gaussian distribution used in this article is replaced with the sub-exponential distribution with a larger range, the relevant issues merit additional investigation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Bickel, P.J. and Levina, E. (2008) Covariance Regularization by Thresholding. *The Annals of Statistics*, **36**, 2577-2604. <https://doi.org/10.1214/08-AOS600>
- [2] Cai, T.T. and Zhou, H.H. (2012) Optimal Rates of Convergence for Sparse Covariance Matrix Estimation. *The Annals of Statistics*, **40**, 2389-2420. <https://doi.org/10.1214/12-AOS998>
- [3] Cai, T.T. and Zhou, H.H. (2012) Minimax Estimation of Large Covariance Matrices under L1-Norm. *Statistica Sinica*, **22**, 1319-1349. <https://doi.org/10.5705/ss.2010.253>
- [4] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <http://www.jstor.org/stable/3085904>
<https://doi.org/10.1198/016214501753382273>
- [5] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. <https://doi.org/10.1198/016214506000000735>
- [6] Rothman, A.J., Levina, E. and Zhu, J. (2009) Generalized Thresholding of Large Covariance Matrices. *Journal of the American Statistical Association*, **104**, 177-186. <https://doi.org/10.1198/jasa.2009.0101>
- [7] Cai, T.T. and Liu, W. (2011) Adaptive Thresholding for Sparse Covariance Matrix Estimation. *Journal of the American Statistical Association*, **106**, 672-684.

-
- <http://www.jstor.org/stable/41416401>
<https://doi.org/10.1198/jasa.2011.tm10560>
- [8] Cai, T.T., Liu, W. and Zhou, H.H. (2016) Estimating Sparse Precision Matrix: Optimal Rates of Convergence and Adaptive Estimation. *The Annals of Statistics*, **44**, 455-488. <https://doi.org/10.1214/13-AOS1171>
- [9] Bickel, P.J. and Levina, E. (2008) Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics*, **36**, 199-227. <https://doi.org/10.1214/009053607000000758>
- [10] Cai, T.T., Zhang, C.H. and Zhou, H.H. (2010) Optimal Rates of Convergence for Covariance Matrix Estimation. *The Annals of Statistics*, **38**, 2118-2144. <https://doi.org/10.1214/09-AOS752>
- [11] Cai, T.T. and Zhang, A. (2016) Minimax Rate-Optimal Estimation of High-Dimensional Covariance Matrices with Incomplete Data. *Journal of Multivariate Analysis*, **150**, 55-74. <https://doi.org/10.1016/j.jmva.2016.05.002>
- [12] Qi, X. (2022) Low Rank Matrix Perturbation Analysis and Estimation for Two Classes of Sparse Covariance Matrices. Ph.D. Thesis, Beijing University, Beijing.
- [13] Shi, W. (2022) Optimal Estimation of Bandable Covariance Matrices Based on Noised Consored Data. MSc. Thesis, Beijing University, Beijing.
- [14] Tsybakov, A.B. (2009) Introduction to Nonparametric Estimation. Springer-Verlag, New York. <https://doi.org/10.1007/b13794>